

AMENDMENT TO THE CLAIMS

1. (Currently Amended) A middleware layer of computer-readable instructions embedded on a computer-readable medium, the instructions being configured to, when executed, facilitate communication between a speech-related application and a speech-related engine, the middleware layer comprising:

a speech component having an application-independent interface configured to be coupled to the application and an engine-independent interface configured to be coupled to the engine and at least one processing component configured to perform speech related services for the application and the engine, wherein the application-independent interface and the engine-independent interface are different interfaces.

2. (original) The middleware layer of claim 1 wherein the speech component includes a plurality of processing components associated with a plurality of different processes, and wherein the speech component further comprises:

a marshaling component, configured to access at least one processing component in each process and to marshal information transfer among the processes.

3. (original) The middleware layer of claim 1 wherein the speech component has an interface configured to be coupled to an audio device, and wherein the speech component further comprises:

a format negotiation component configured to negotiate a data format of data used by the audio device and data used by the engine.

4. (original) The middleware layer of claim 3 wherein the format negotiation component is configured to reconfigure the audio

device to change the data format of the data used by the audio device.

5. (original) The middleware layer of claim 3 wherein the format negotiation component is configured to reconfigure the engine to change the data format of the data used by the engine.

6. (original) The middleware layer of claim 3 wherein the format negotiation component is configured to invoke a format converter to convert the data format of data between the engine and the audio device to a desired format based on the data format used by the audio device and the data format used by the engine.

7. (original) The middleware layer of claim 1 wherein the processing component comprises:

a lexicon container object configured to contain a plurality of lexicons and to provide a lexicon interface to the engine to represent the plurality of lexicons as a single lexicon to the engine.

8. (original) The middleware layer of claim 7 wherein the lexicon container object is configured to, once instantiated, load one or more user lexicons and one or more application lexicons from a lexicon data store.

9. (original) The middleware layer of claim 8 wherein the lexicon interface is configured to be invoked by the engine to add a lexicon provided by the engine.

10. (original) The middleware layer of claim 1 wherein the processing component comprises:

a site object having an interface configured to receive

result information from the engine.

11. (original) The middleware layer of claim 1 wherein the engine comprises a text-to-speech (TTS) engine and wherein the processing component comprises:  
a first object having an application interface and an engine interface.
12. (original) The middleware layer of claim 11 wherein the application interface exposes a method configured to receive engine attributes from the application and instantiate a specific engine based on the engine attributes received.
13. (original) The middleware layer of claim 11 wherein the application interface exposes a method configured to receive audio device attributes from the application and instantiate a specific audio device based on the audio device attributes received.
14. (original) The middleware layer of claim 11 wherein the first object includes a parser configured to receive input data to be synthesized and parse the input data into text fragments.
15. (original) The middleware layer of claim 11 wherein the engine interface is configured to call a method exposed by the engine to begin synthesis.
16. (original) The middleware layer of claim 1 wherein the engine comprises a speech recognition (SR) engine and wherein the processing component comprises:  
a first object having an application interface and an engine interface.
17. (original) The middleware layer of claim 16 wherein the

application interface exposes a method configured to receive recognition attributes from the application and instantiate a specific speech recognition engine based on the engine attributes received.

18. (original) The middleware layer of claim 16 wherein the application interface exposes a method configured to receive audio device attributes from the application and instantiate a specific audio device based on the audio device attributes received.

19. (original) The middleware layer of claim 16 wherein the application interface exposes a method configured to receive an alternate request from the application and to configure the speech component to retain alternates provided by the SR engine for transmission to the application based on the alternate request.

20. (original) The middleware layer of claim 16 wherein the application interface exposes a method configured to receive an audio information request from the application and to configure the speech component to retain audio information recognized by the SR engine based on the audio information request.

21. (original) The middleware layer of claim 16 wherein the application interface exposes a method configured to receive bookmark information from the application identifying a position in an input data stream being recognized and to notify the application when the SR engine reaches the identified position.

22. (original) The middleware of claim 16 wherein the engine interface is configured to call the SR engine to set acoustic profile information in the SR engine.

23. (original) The middleware of claim 16 wherein the engine

interface is configured to call the SR engine to load a grammar in the SR engine.

24. (original) The middleware of claim 16 wherein the engine interface is configured to call the SR engine to load a language model in the SR engine.

25. (original) The middleware layer of claim 16 wherein the application interface exposes a method configured to receive a grammar request from the application and to instantiate a grammar object based on the grammar request.

26. (original) The middleware layer of claim 25 wherein the grammar object includes a word sequence data buffer and an interface configured to provide the SR engine with access to the word sequence data buffer.

27. (original) The middleware layer of claim 25 wherein the grammar object includes a grammar to be used by the SR engine.

28. (original) The middleware layer of claim 27 wherein the grammar includes words, rules and transitions and wherein the grammar object includes an application interface and an engine interface.

29. (original) The middleware layer of claim 28 wherein the application interface exposes a grammar configuration method configured to receive grammar configuration information from the application and configure the grammar based on the grammar configuration information.

30. (original) The middleware layer of claim 29 wherein the grammar configuration method is configured to receive rule

activation information and activate or deactivate rules in the grammar based on the rule activation information.

31. (original) The middleware layer of claim 29 wherein the grammar configuration method is configured to receive grammar activation information and enable or disable grammars in the grammar object based on the grammar activation information.

32. (original) The middleware layer of claim 29 wherein the grammar configuration method is configured to receive word change data, rule change data and transition change data and change words, rules and transitions in the grammar in the grammar object based on the grammar received data.

33. (original) The middleware layer of claim 28 wherein the engine interface is configured to call the SR engine to load the grammar in the SR engine.

34. (original) The middleware layer of claim 33 wherein the engine interface is configured to call the SR engine to update a configuration of the grammar in the SR engine.

35. (original) The middleware layer of claim 33 wherein the engine interface is configured to call the SR engine to update an activation state of the grammar in the SR engine.

36. (previously presented) The middleware layer of claim 1 wherein the processing component further comprises:

a site object exposing an engine interface configured to receive information from the speech-related engine.

37. (original) The middleware layer of claim 36 wherein the engine interface on the site object is configured to receive

result information from the SR engine indicative of recognized speech.

38. (original) The middleware layer of claim 36 wherein the engine interface on the site object is configured to receive update information from the SR engine indicative of a current position of the SR engine in an audio input stream to be recognized.

39. (original) The middleware layer of claim 36 wherein the processing component further comprises:

a result object configured to obtain the result information from the site object and expose an interface configured to pass the result information to the application.

Claims 40-46 (Previously Canceled)

47. (Currently Amended) A multi-voice speech synthesis middleware layer of computer-readable instructions embedded on a computer-readable medium, the instructions being configured to, when executed, facilitate communication between one or more applications and a plurality of text-to-speech (TTS) engines, the multi-voice speech synthesis middleware layer comprising:

at least a first voice object having an application interface configured to receive TTS engine attribute information from the application and to instantiate first and second TTS engines based on the TTS attribute information, to receive a speak request requesting at least one of the TTS engines to speak a message, and to receive priority information associated with each speak request indicative of a precedence each speak request is to take;

wherein the first voice object has an engine interface

configured to call a specified one of the first and second TTS engines to synthesize input data;  
wherein the at least first voice object is configured to receive a normal priority associated with a message and to call the TTS engines so the message with normal priority is spoken in turn; and  
wherein the at least first voice object is configured to receive a speakover priority associated with a message and to call the TTS engines so the message with speakover priority is spoken at a same time as other currently speaking messages.

Claims 48 and 49 (Previously Canceled)

50. (Previously Presented) The multi-voice speech synthesis middleware layer of claim 47 wherein the at least first voice object is configured to receive an alert priority associated with a message and to call the TTS engines so the message with alert priority is spoken with precedence over messages with normal and speakover priority.

Claims 51 and 52 (Presently Canceled)

53. (original) A method of formatting data for use by a speech engine and an audio device, comprising  
obtaining, at a middleware layer which facilitates communication between the speech engine and an application, a data format for data used by the engine;  
obtaining, at the middleware layer, a data format of data used by the audio device;  
determining, at the middleware layer, whether the engine data format and the audio data format are consistent; and  
if not, utilizing the middleware layer to attempt to change



the data format of the data used by at least one of the engine and the audio device.

54. (original) The method of claim 53 and further comprising:  
if the attempt to change the data format used by the at least one of the engine and the audio device is unsuccessful, invoking a format converter to change data format for data between the engine and the audio device to ensure the data formats are consistent.